

A microscopic view of several chromosomes, appearing as thick, purple, X-shaped structures against a blue background. The chromosomes are in various orientations, with some in sharp focus and others blurred in the background.

Análisis de BRCA1

Bioinformática

Ramiro Olivera Fedi
Julián Antonielli

El cáncer de mama afecta a *una de cada*
ocho mujeres durante su vida.

El cáncer de mama afecta a *una de cada ocho mujeres* durante su vida.

Se estima que *más de 40.000* mujeres morirán este año por esta enfermedad.

> BRCA1

Gen que codifica una proteína que trabaja en el mantenimiento de la estabilidad genómica, *actúa como un supresor tumoral*, y juega un papel en la transcripción, reparación del ADN y su recombinación.

> BRCA1

Gen que codifica una proteína que trabaja en el mantenimiento de la estabilidad genómica, *actúa como un supresor tumoral*, y juega un papel en la transcripción, reparación del ADN y su recombinación.

Las *mutaciones en este gen* son responsables de cerca del 40% de los cánceres de mama heredados, y más del 80% entre cáncer de mama y ovario heredados.

mRNA a FASTA

Se leyó una secuencia de nucleótidos del mRNA seleccionado (NM_007294) del gen BRCA1 en formato GenBank y traducirlos a las posibles secuencias de aminoácidos, finalmente guardandolos en distintos archivos de formato FASTA.

Reading frame

En las regiones donde los codones-stop son muy frecuentes, en general, no puede traducirse ningún polipéptido que tenga alguna funcionalidad específica dada la baja cantidad de aminoácidos por cadena proteica producida. En el *reading frame* 2, hay *una región de alrededor de 1800 aminoácidos consecutivos* a la cual se le podría llegar a asignar una funcionalidad. En contraste, los otros reading frames se encuentran interrumpidos por una alta frecuencia de codones-stop a lo largo de la secuencia.

Reading frame

KYIGTERNGPICSSK*KSIRCH*CTAENERSHESGVDQGT
M*E*YNQKEPTRKYEI*STC*RAIENHLCFSA*HRFGVCKQL
CQKTSTE*TRKSFLAGNQSQCPTL*PWNCENSEDKAADTTSK
IVTNHPSRNQG*NQFGFCKKGCL*IF*DGCNKY*TSSTQ***
MWHKYSCQLITA*EQQFITH*RQNECRKG*IL**KQTAWLSK
C*SPV*EKRME*AETAMLRES*RY*RCSLDNTK*QHSES**V
GRSK*GR*IFWFFRENRLTGQ*SS*GFNM*K*KSSLQISRE*
RSICY*ATDNTRASPHK*IKA*KETYIRPSS*GFYQESRFGS
K*KNR*EYSE*EKS*DNRTTRKDTGEONES*TVKQYKQYK


reading frame 1

GRNDI*TRKSEQES*TRFSQ*VELEERIG*QED*QLEETANS*NTFA
AKRLLQSEPENPSLQETSLSVQLSNLGTVRTLR TKQRIQPQKTSVY
LLQITPQGTRDEISLDSAKKAACEFSETDVTNTEHHQPSNNDLNTT
CGTNTHASSLQHENSLLLT KDRMNVEKA EFCNKSQKQPLARSQHN
ADPLCERKEWNKQKLPCSENPRDTE DVPWITLNSSI QKVNEWFSRS
DVLNEVDEYSGSSEKIDLLASDPHEALICKSERVHKS SVESNIEDK
GAFVTEPQIIQERPLTNKLKRKRRTSGLHPEDFIKKADLAVQKTP
ENKTKGDSIQNEKNPNPIESLEKESAFKTKA EPISSSISNMELELN
ELVVSRLNLPNCTELQIDSCSSSEEIKKKKYNQMPVRHSRNLQLM
SDTEDELKLTNAPGSETKCSNTSELKEEVNPSLDPFEKFEKLETVK

reading frame 2

BLAST

Se generaron reportes de BLAST remoto para cada posible reading frame. Si nos basamos en los resultados del BLAST es fácil notar que el *reading frame* correcto es el segundo. Solo el hecho que el *bit score* del primer resultado de la query hecha con ese *reading frame* sea de más de 3500, mientras que el siguiente mejor con otro reading frame es de 114 es evidencia suficiente para probarlo.



Alineamiento

Con los resultados obtenidos del BLAST de la secuencia de aminoácidos con el *reference frame* correcto, se seleccionaron 3 especies a comparar: *Pan paniscus*, *Pan troglodytes* y *Pongo abelii*.

Especie	mRNA
Homo Sapiens	NM_007294
Pan paniscus	NP_001288687
Pan troglodytes	XP_016785970
Pongo abelii	XP_003778880

Alineamiento

```
XP_003778880.2 -----
. yldfvfeaaarplgfrngkarelqinncdcaaardfldggqavgfllrgpcaqeaftlc
XP_016785970.2 -----
NP_001288687.1 -----

XP_003778880.2 -----MDLSAvRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKL
. sgssleqkeMDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKL
XP_016785970.2 -----MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKL
NP_001288687.1 -----MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKL

XP_003778880.2 LNQKKGPSQCPLCKNDITKRSIQESTRFSQIqVEELLKIIICAFQqLDTGLEYANSYNFAKKE
. LNQKKGPSQCPLCKNDITKRSIQESTRFSQIqVEELLKIIICAFQqLDTGLEYANSYNFAKKE
XP_016785970.2 LNQKKGPSQCPLCKNDITKRSIQESTRFSQIqVEELLKIIICAFQqLDTGLEYANSYNFAKKE
NP_001288687.1 LNQKKGPSQCPLCKNDITKRSIQESTRFSQIqVEELLKIIICAFQqLDTGLEYANSYNFAKKE

XP_003778880.2 NNSPEHLKDEVSIIQSMGYRNRARLLQSEPENPSLQETSpSVQLSNLGTVRTLRTKQRI
. NNSPEHLKDEVSIIQSMGYRNRARLLQSEPENPSLQETSpSVQLSNLGTVRTLRTKQRI
XP_016785970.2 NNSPEHLKDEVSIIQSMGYRNRARLLQSEPENPSLQETSpSVQLSNLGTVRTLRTKQRI
NP_001288687.1 NNSPEHLKDEVSIIQSMGYRNRARLLQSEPENPSLQETSpSVQLSNLGTVRTLRTKQRI
```

Salida del alineamiento múltiple con MUSCLE

Alineamiento

Todas las secuencias coinciden en gran parte.

Al tratarse de un gen regulador del ciclo celular, se trata de un mecanismo muy común conservado a lo largo de las distintas especies. Aún más en especies similares como las elegidas. Por lo que podemos concluir que la estructura de la proteína es la misma (o muy similar).

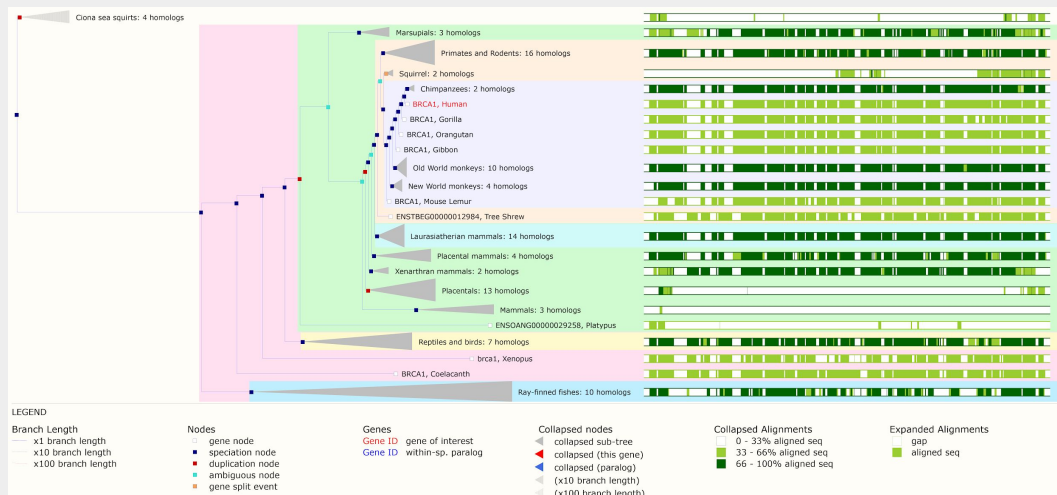
Finalmente, podemos concluir que los diferentes aminoácidos presentes en la mismas ubicaciones no alteran la estructura de la proteína, es decir, las mutaciones de estos fueron inocuas a su correcto funcionamiento.

Genes homólogos

Con la base de `Ensembl` en la cual se obtuvieron 103 genes homólogos, visualizados a través del Gene Tree, en el cual se puede ver con más detalle las especies y las familias en la cuales está identificado algún gen homólogo al `BRCA1`. `HomoloGene` muestra solo 8.



Genes homologs



BRCA1, *H.sapiens*
 breast cancer 1, early onset
 BRCA1, *P.troglodytes*
 breast cancer 1, early onset
 BRCA1, *M.mulatta*
 breast cancer 1, early onset
 BRCA1, *C.lupus*
 breast cancer 1, early onset
 BRCA1, *B.taurus*
 breast cancer 1, early onset
 Brca1, *M.musculus*
 breast cancer 1
 Brca1, *R.norvegicus*
 breast cancer 1, early onset
 BRCA1, *G.gallus*
 breast cancer 1, early onset

Genes homólogos

Dado que se trata de un gen que genera supresores de tumores, y altamente relacionado con la reproducción celular, entendemos que *es un gen que se encontrará en especies variadas, incluso de distintas familias*. Por eso podemos ver que se encuentran desde mamíferos hasta reptiles y marsupiales.

Transcriptos

En búsqueda de la distinta cantidad de transcritos y formas alternativas de splicing para BRCA1, se realizó una búsqueda en Ensembl, en el que se encontraron los siguientes 33 transcritos, de los cuales solo se expresan 26.

Las diferencias entre los tamaños de las proteínas nos sugiere que *la estructura será distinta y por lo tanto sus respectivas funciones diferentes*. Además, al estar identificadas por distintos nombres, se nos sugiere que están involucradas en distintos mecanismos.

Transcripts

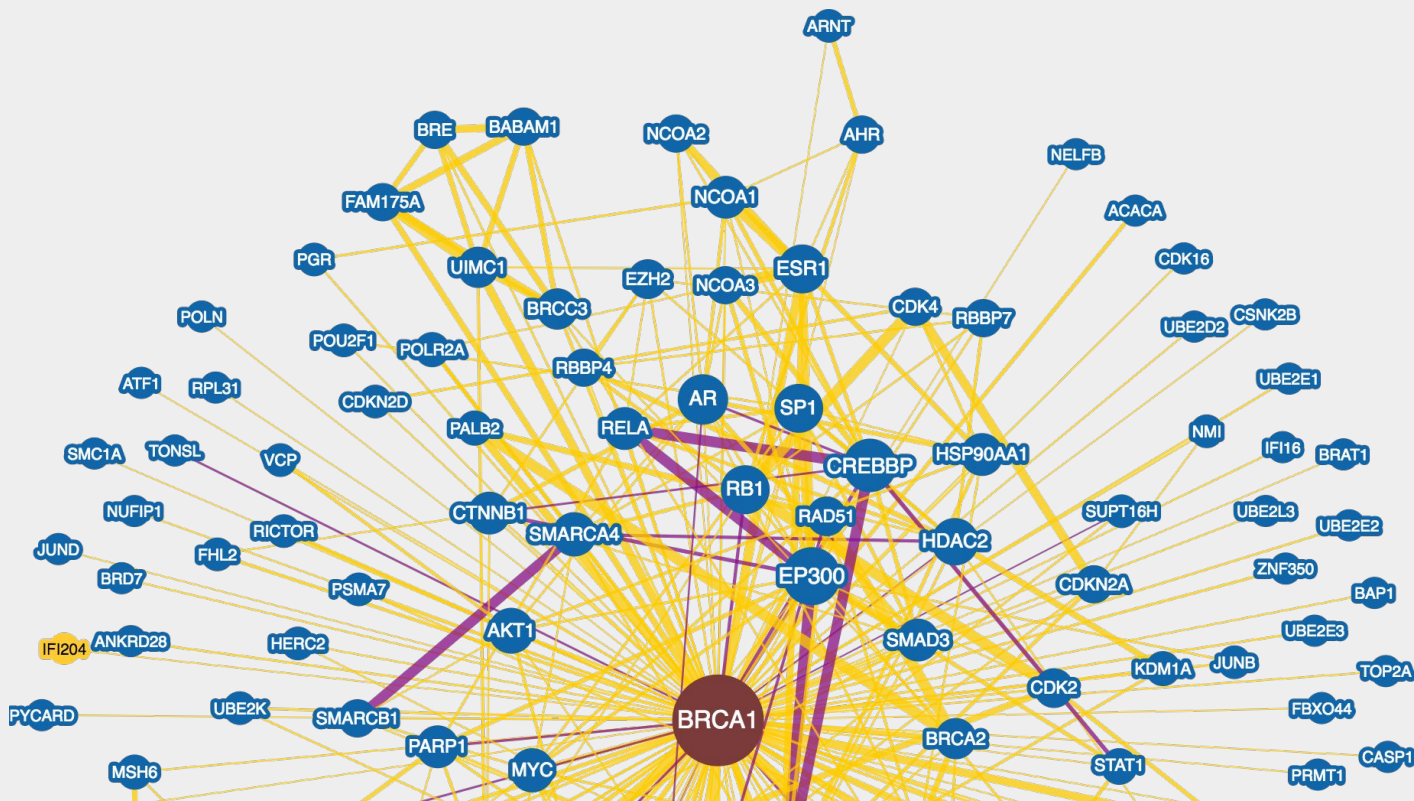
Show/hide columns										Filter			
Name	Transcript ID	bp	Protein	Translation ID	Biotype	CCDS	UniProt	RefSeq	Flags				
BRCA1-210	ENST00000471181.7	7270	1884aa	ENSP00000418960	Protein coding	CCDS11456	P38398	NM_007300 NP_009231	TSL:1	GENCODE basic	APPRIS P4		
BRCA1-203	ENST00000357654.8	7207	1863aa	ENSP00000350283	Protein coding	CCDS11453	P38398	NM_007294 NP_009225	TSL:1	GENCODE basic	APPRIS ALT2		
BRCA1-221	ENST00000493795.5	5732	1816aa	ENSP00000418775	Protein coding	CCDS11459	P38398	NM_007297 NP_009228		TSL:5	GENCODE basic		
BRCA1-208	ENST00000468300.5	3273	699aa	ENSP00000417148	Protein coding	CCDS11455	P38398	NM_007299 NP_009230		TSL:1	GENCODE basic		
BRCA1-219	ENST00000491747.6	2379	759aa	ENSP00000420705	Protein coding	CCDS11454	A0A024R1V0 P38398	NM_007298 NP_009229		TSL:5	GENCODE basic		
BRCA1-202	ENST00000354071.7	4497	1399aa	ENSP00000326002	Protein coding	-	Q5YLB2	-		TSL:1	GENCODE basic		
BRCA1-201	ENST00000352993.7	3668	721aa	ENSP00000312236	Protein coding	-	A0A024R1Z8 P38398	-		TSL:5	GENCODE basic		
BRCA1-232	ENST00000644379.1	2571	659aa	ENSP00000496570	Protein coding	-	-	-		CDS 5' incomplete			
BRCA1-230	ENST00000634433.1	2534	798aa	ENSP00000489431	Protein coding	-	A0A0U1RRA9	-		CDS 3' incomplete	TSL:5		
BRCA1-209	ENST00000470026.5	2108	649aa	ENSP00000419274	Protein coding	-	E7EWN5	-		CDS 3' incomplete	TSL:1		
BRCA1-214	ENST00000477152.5	1980	622aa	ENSP00000419988	Protein coding	-	E9PH68	-		CDS 3' incomplete	TSL:1		
BRCA1-215	ENST00000478531.5	1972	623aa	ENSP00000420412	Protein coding	-	E7EUM2	-		CDS 3' incomplete	TSL:1		
BRCA1-222	ENST00000493919.5	1948	572aa	ENSP00000418819	Protein coding	-	B7ZA85	-		CDS 3' incomplete	TSL:2		
BRCA1-223	ENST00000494123.5	1612	473aa	ENSP00000419103	Protein coding	-	Q3B891	-		CDS 3' incomplete	TSL:1		
BRCA1-216	ENST00000484087.5	1495	498aa	ENSP00000419481	Protein coding	-	H0Y8B8	-		CDS 5' and 3' incomplete		TSL:5	
BRCA1-226	ENST00000591534.5	1282	354aa	ENSP00000467329	Protein coding	-	K7EPC7	-		TSL:1	GENCODE basic		

Interacciones

Utilizando NCBI podemos encontrar 1150 proteínas que interactúan con el gen BRCA1, y utilizando Uniprot, encontramos que la base de datos BioGrid logra identificar alrededor de 1000.

Debido a la gran cantidad de interacciones no es fácil identificar un patrón entre las mismas. No obstante, analizando algunas de las proteínas que interactúan con BRCA1 podemos notar que la mayoría están asociadas a mecanismos de reparación o identificación de daño genético. Lo que nos lleva a confirmar que dicha proteína está asociada a *pathways* de reparación genética.

Interacciones



Interacciones

Analizando un caso particular, podemos ver la interacción de `BRCA1` con la proteína `RAD51`. Ésta última se encarga de asistir en la reparación de daños de cadena doble en el ADN, y puede encontrarse en la mayoría de las células eucariotas, desde levadura hasta células humanas.

En cuanto a la relación que existe entre `BRCA1` y `RAD51`, encontramos que la primera regula la segunda en respuesta al daño en el ADN con implicancias en la cohesión de cromátidas hermanas, estabilidad genómica y en la carcinogénesis.

Componentes celulares

Utilizando la base de datos de genes humanos GeneCards, buscamos información sobre los componentes celulares que forman parte de BRCA1.

GO ID	Qualified GO term	Evidence	PubMed IDs
GO:0000151	ubiquitin ligase complex	NAS	14976165
GO:0000793	condensed chromosome	IEA	
GO:0000794	condensed nuclear chromosome	IEA	
GO:0000800	lateral element	IDA	9774970
GO:0005634	nucleus	IEA,IDA	17525340
GO:0005654	nucleoplasm	TAS	
GO:0005694	chromosome	ISS	
GO:0005737	cytoplasm	IEA,IDA	20160719
GO:0005886	plasma membrane	IDA	21282464
GO:0008274	gamma-tubulin ring complex	NAS	12214252
GO:0016020	membrane	IEA	
GO:0016021	integral component of membrane	IEA	

Procesos biológicos

Utilizando el mismo sitio encontramos 68 procesos biológicos a los que pertenece BRCA1.

GO ID	Qualified GO term	Evidence	PubMed IDs
GO:0000724	double-strand break repair via homologous recombination	IDA	17349954
GO:0000729	DNA double-strand break processing	TAS	
GO:0000731	DNA synthesis involved in DNA repair	TAS	
GO:0000732	strand displacement	TAS	
GO:0006260	DNA replication	TAS	
GO:0006281	DNA repair	IEA	
GO:0006301	postreplication repair	IDA	17349954
GO:0006302	double-strand break repair	IMP,IDA	17525340
GO:0006303	double-strand break repair via nonhomologous end joining	TAS	
GO:0006310	DNA recombination	IEA	
GO:0006349	regulation of gene expression by genetic imprinting	IEA	
GO:0006351	transcription, DNA-templated	IEA	

Funciones moleculares

Continuando con el análisis de los resultados, encontramos que esta proteína trabaja en las siguientes 16 funciones moleculares

GO ID	Qualified GO term	Evidence	PubMed IDs
GO:0001105	RNA polymerase II transcription coactivator activity	IMP	9662397
GO:0003677	DNA binding	TAS,IEA	9662397
GO:0003684	damaged DNA binding	IEA	
GO:0003713	transcription coactivator activity	NAS	15572661
GO:0003723	RNA binding	IDA	12419249
GO:0004842	ubiquitin-protein transferase activity	IDA,IEA	12890688
GO:0005515	protein binding	IPI	8944023
GO:0008270	zinc ion binding	IEA	
GO:0015631	tubulin binding	NAS	12214252
GO:0016740	transferase activity	IEA	
GO:0019899	enzyme binding	IPI	15965487
GO:0021625	histone H3 binding	IPI	17070005

Variaciones genéticas


Como se discutió hasta ahora `BRCA1` es un gen supresor de tumores. Para los individuos con variantes de `BRCA1` causales por ejemplo, la `OMIM` resume los siguientes valores:

- El riesgo de cáncer de mama a lo largo de la vida para portadores de mutaciones es de entre el 80% y el 90%
- El riesgo de cáncer de ovario a lo largo de la vida para portadores de mutaciones es de entre el 40% y el 50%
- Aumenta el riesgo de cáncer de mama bilateral

Variaciones genéticas

Algunas de las mutaciones menos raras, causales del cáncer de mama son `rs386833395` y `rs80357906`, mejor conocidas como `185delAG` y `5382insC`.

Dichas variantes afectan principalmente al grupo étnico conocido como *Judíos Asquenazí*. Esta denominación se refiere a los judíos que se asentaron en Europa Central y Oriental. Los asquenazíes son los descendientes de las comunidades judías medievales establecidas a lo largo del Río Rin.



The background of the image is a microscopic view, likely a scanning electron micrograph, showing various elongated, rod-like structures. Some of these structures are a vibrant purple color, while others are a darker, muted blue. They appear to have a textured, almost granular surface. The structures are scattered across the frame, with some in sharp focus and others blurred in the background, creating a sense of depth. The overall lighting is soft, highlighting the three-dimensional nature of the objects.

MUCHAS GRACIAS.